'apparatus' in Goldenthal et al that corresponds to the 'apparatus' of claim 2? Applicants are at a loss here because the Examiner did not specify any correspondence.

In the context of the FIG. 1 embodiment, the apparatus might be

(a) the entirety of the system described in the FIG., with the 'transmission' referring to the sound that is applied to microphone 110,

(b) that portion of the system that excludes microphone 110, with the 'transmission' referring to the signal flowing from microphone 110 to ADL 120, or

(c) rendering system 240.

The FIG. 3 embodiment yields, effectively, option (b) above.

Claim 2, however, specifies a method where a first signal is generated that is a stream of signals for generating sounds, a second signal of commands to a video synthesizer in the apparatus, and a combining of the first and second signals to form the signal that is being transmitted to the apparatus. Clearly, options (a) and (b) do not apply because only an audio signal is being transmitted to the microphone, or from the microphone. That leaves only option (c), where the "apparatus" to which a combined signal is sent is rendering system 240. However, it is quite clear from FIG. 1, and from the text of the reference, that no such combined signal is created, or transmitted, to rendering system 240. Therefore, it is respectfully submitted that claim 2 is not anticipated by the reference.

Moreover, since the visemes are that are produced and outputted on line 116 are already in synchronism with the speech (by virtue of the processing by elements 200, 113, 130, and 115) there is no incentive to combine the signals of line 116 and the output line of element 117 – only to be separated again within rendering system 240.

Therefore, it is believed that claim 2 is neither anticipated nor rendered obvious by Goldenthal et al. Since claim 2 is not anticipated or rendered obvious by Goldenthal et al, it is respectfully submitted that claims 3-11, which depend on claim 2, are also not anticipated or rendered obvious by Goldenthal et al. At least some of them also posses other limitations that make them neither anticipated nor rendered obvious by Goldenthal et al.

Claim 3 specifies that the commands transmitted in the claim 2 method are FAP signals, and there is no mention or suggestion of such signals anywhere in the Goldenthal

et al reference. As to what those signals are, the Examiner's attention is respectfully directed to page 1, lines 21-28, where it states:

> One of these AVOs is the Face Object, which allows animation of synthetic faces, sometimes called Talking Heads. It consists of a 3D synthetic visual object representing a human face, a synthetic audio object, and some additional information required for the animation of the face. Such a scene can be defined using the BInary Format for Scene (BIFS), which is a language that allows composition of 2D and 3D objects, as well as animation of the objects and their properties.
>
> The face model is defined by BIFS through the use of nodes. The Face Animation Parameter node (FAP) defines the part of the face has to be animated.

This, claim 3 is not anticipated or rendered obvious by Goldenthal et al. The above remarks apply with equal vigor to claims 34 and 37.

Claim 4 specifies that the FAP signals include prosody and timing. Though timing is mentioned in the Goldenthal et al reference, it is not mentioned in connection with FAP signals. Prosody is not mentioned at all.

Claim 5 specifies that the FAP signals exclude signals that provide viseme information. Aside from the fact that FAP signals are simply not mentioned in Goldenthal et al, the Examiner's explanation for rejecting the claim is that Goldenthal et al "teaches other image information that can include non-visemes." Respectfully, that does not support a rejection, since an assertion that something can include items other than items A is not the same as a specification that something is *devoid* of items A.

As for claim 9, it specifies that the first signal – that is the signal which includes signals for generating sounds – contains text. In support of the claim's rejection, the Examiner cited col. 4, line 65 – col. 5, line 4. However, the cited passages states:

> It should be understood that the invention can also be used to synchronize visual images to streamed audio signals in real time. For example, a web-based "chat room" can be configured to allow multiple users to concurrently participate in a conversation with multiple synchronized talking heads. The system can also allow two client computers to exchange audio messages directly with each other.

There is no mention of text in the passage. Accordingly, it is believed that claim 9 is neither anticipated nor rendered obvious by Goldenthal et al. This above remarks apply with equal vigor to claims 32 and 33.

Claim 10 specifies that, in addition to the first signal stream and the second signal stream, there is a "stream of parameter signals for said video synthesizer." In the Goldenthal et al system, however, there is no signal created or applied to element 240 that corresponds to this additional stream. The facial images that the Examiner mentions in connection with this claim are mentioned in the general statement "[A]lthough the example system 1000 is described in terms of human speech and facial images, it should be understood..." The reference to 'human speech' relates merely to the signal outputted by element 117, and the reference to 'facial images' relates merely to the signal outputted by element 115. It is respectfully submitted, therefore, that claim 10 is neither anticipated nor rendered obvious by Goldenthal et al. This above remarks apply with equal vigor to claims 36-40.

Claim 11 narrows the definition of claim 10 by specifying that this additional stream comprises "face model information." No face model is mentioned in Goldenthal et al.

Claim 12 is an independent apparatus claim (in contrast to all of the previously discussed claims, which are method claims). It specifies a decoder, a converter for developing sound, and a video synthesizer for developing images. In the Goldenthal et al system, it is rendering system 240 that performs the functions of the converter and the video synthesizer. Leaving for a moment the question of whether the converter and video synthesizer elements specified by claim 12 are anticipated by Goldenthal et al, it is clear that the decoder of claim 12 has to correspond to one or more of the other elements in the Goldenthal et al system, if the Examiner's rejection can be sustained.

It is respectfully submitted that the Examiner's rejection cannot be sustained. The decoder defined in claim 12 is one that is responsive to an input signal comprising signals representing audio and embedded video synthesis command signals. None of the elements in the Goldenthal et al system are responsive to such a signal. Further, the decoder defined in claim 12 is one that separates the video synthesis command signals stream from signals representing the audio signal stream. No such separating is performed by any of the elements in the Goldenthal et al system. Accordingly, it is respectfully submitted that claim 12 is not anticipated by Goldenthal et al, and nor is it obvious in view of Goldenthal et al. Since claim 12 is neither anticipated nor rendered

obvious by Goldenthal et al, it is respectfully submitted that claims that depend on claim 12 (26-28) also are neither anticipated nor rendered obvious by Goldenthal et al. Moreover, at least some of them also posses other limitations that make them neither anticipated nor rendered obvious by Goldenthal et al.

Regarding claim 26, it specifies that the decoder generates additional signals that interpolate between the separated command signals from the input signal. No such generation is taught by Goldenthal et al. The Examiner asserted that in col. 4, lines 14-24, Goldenthal et al teach "pairing commands generated from acoustic phonetic units translated to visemes." Respectively, that is not the case. The cited passage mentions no pairing of anything. It merely teaches that the acoustic-phonetic units are formatted as data records, each having a start time field, an end time field, and an identification. These units are translated to visemes. There is no mention in this passage of an input signal with embedded video synthesis command signals, and there is no mention of the creation of additional video synthesis command signals.

Regarding claim 27, the Examiner asserted that Goldenthal et al teach interposing signals, pointing to sub-blocks 130 and 131 of FIG. 1. Respectfully, applicants do not agree. Elements 130 and 131 are hardware elements; not signals. Further, to the extent that elements 130 and 131 are "interposed," they are interposed between elements 113 and 115, which are also hardware elements; not signals. Claim 27, in contradistinction, speaks of additional command signals, and specifies "where each set of said additional command signals that are interposed between a pair of command signals interpolates between said pair of command signals."

Rejected claim 31 is an independent method claim. The Examiner has made no specific comments regarding this rejection of the claim.

Claim 31 specifies a step of receiving an input signal that comprises signals representing audio and embedded video synthesis command signals. No such input signal is taught in Goldenthal et al and, therefore, this step is not anticipated by Goldenthal et al. The second step of the claim specifies a step of separating the input signal into two streams. Since no such input signal exists in Goldenthal et al, it is not surprising that no such step is taught or suggested by Goldenthal et al. Accordingly, it is respectfully submitted that claim 31 is neither anticipated nor rendered obvious by

Goldenthal et al. Since claim 31 is neither anticipated nor rendered obvious by Goldenthal et al, it follows that the claims which depend on claim 31 (32, 34-40) are also neither anticipated nor rendered obvious by Goldenthal et al.

Claims 2, and 12-25 were rejected under 35 USC 102(b) as being anticipated by Gasper, US Patent 4,8884,972. Applicants respectfully traverse.

Gasper describes a system that creates sounds and an image that is synchronized to the sound. However, a close scrutiny of the Gasper teachings reveals that the elements, or modules, described by Gasper are not at all the same as, or similar to, the elements defined by the rejected claims.

Basically, Gasper describes a system that operates in a different manner, and with structure that is different from that which applicants claim. Unfortunately, to demonstrate this fact, it appears necessary to describe the Gasper system is some detail, because this system is fairly complicated.

Gasper describes a system where a user is presented with tile images on a computer screen. The screen also has a simulated "talking head." Each tile has an imprinted image of a letter or a phonogram (a set of letters that produce a sound). The objective of the system is to allow a user to move tiles, to concatenate them to form words, and to direct the system to sound out the formed words. While the words are sounded out, the talking head image is changed in synchronism with the created sound so as to give the impression that the created sound comes from, or pronounced by, the talking head's "mouth." Gasper calls the talking head a "synactor."

Once a user is finished concatenating tiles and is ready for the computer to animate (both in sound and in the synactor image) the created word, a script is created. When the script is ready, it is applied to processor 51 which generates the orthophonetic sounds (sounds that correspond to correct speaking), and to processor 53, which generates the proper synactor images.

Using information derived from a program that contains characteristics of the synactor, from synactor behavior controller 49, or controller 44, or tile controller 37, a central controller 10 of the Gasper system writes the orthophonetic animation script for processor 51, and the synactor animation script for processor 53.

More specifically, an input text (e.g., the word to be sounded out and imaged by the synactor) is applied to text-to-phoneme translator 40, which translates the text to a phonetic string. Translator 40 also creates record for correctly speaking the phonetic string (OCREC) from data contained in RAM 20. Thereafter, the OCREC and the phonetic string are applied to encoder 41, which converts the phonetic string to component phonetic codes, and maps the phonetic code string to a phocode representation (e.g. 49, 19, 57.)

The phocodes string is applied to tile controller 37, along with the OCREC record. Controller 37 adds orthographic information (information about the display of the characters) and passes the information to controller 10. Controller 10 accesses RAM 20 for the synactor data structures corresponding to the created phocoded string. This data includes the image sequences and context dependent manipulations as programmed in the RAVEL source code for each phocode for this synactor (see, illustratively, Gasper's FIG. 6).

An application controller 31 and microprocessor 10 then provide information to a behavior controller 49, describing what is going on much as a person's external and internal senses communicate with the brain where certain events or combinations trigger behavioral traits. The behavior controller 49 accesses RAM 20 for data structures to simulate personality and give each synactor a character of its own.

Controller 10 uses the provided information to generate the raw synactor, orthophonetic and audio scripts, and directs audio script generator 42, orthophonetic animation script generator 52 and synactor animation script generator 54 to process the raw scripts to produce final scripts. This includes inserting a rest position at the end of each script, generating inbetweens, etc. Once the final scripts are generated, the scripts are acted out; i.e., coordinated in real-time by the Real Time Coordinator 55 to provide the audio and display the associated time synchronized video called for by the user input.

As indicated above, the script includes "inbetween" image specifications. An inbetween specifies the image to be displayed between two other specified images. For example, with reference to Gasper's FIG. 6 and the element within in that is labeled 638, an inbetween specification "33 5 65 3" means

Beutnagel 4-1-13-3

> For instance the first inbetween statement 638 specifies to the synactor script generator 54 that anytime the image numbered 33 is to be displayed on the screen followed immediately by the image numbered 5, the image number 65 is to be inserted between those two images for a duration of 3 cycles. In this instance the display time allowed for the image numbered 33 is reduced by 3 cycles to provide the display time for the inbetween image number 65. (col. lines 26-28).

A careful study of the Gasper teachings reveals that there is no teaching of what is contained in an OCREC record, or how to create it. There is also no teaching of how to create phocodes, or how to map the phonetic codes to phocodes. The impression left by the description is that it's just a one to one mapping, but such mapping is merely a change in representation.

There is also no teaching of what creates the image sequences that are obtained by controller 10 from RAM 20, how they are created, how the context dependent manipulations are programmed, or how they affect the image sequences.

Further, there is no teaching of what the structures are that RAM 20 provides to behavior controller 49, the conditions under which they are inserted, and how they modify (if at all) the images that are displayed.

What is clear, however, is that the Gasper system does not deal with a model of the talking head, where an image is created by providing parameters to the model.

Claim 1, in contradistinction, specifies a decoder that is responsive to an input signal "comprising text and FAP information" – where FAP was clearly defined in the application, as remarked above.

Consequently, it is respectfully submitted no such decoder is described by Gasper and, in particular, that controller 10 is not responsive to an input signal that contains FAP information.

Moreover, claim 1 specifies that the decoder separates the FAP information from the text. Since there is no FAP information, there certainly is no separation of FAP information from text; and in any event, controller 10 does no separating.

Further, element 10, as the central control element of FIG. 1 and FIG. 3, does not even receive text. As indicated above, the input text is converted in text-to-phoneme translator 40 to a phoneme string, prior to any action by element 10. Indeed, the Examiner would be more correct to equate the decoder of claim 1 to translator 40 rather

than to element 10, although that would still not correspond to the decoder of claim 1 because of the absence of FAP information, and any treatment of FAP information in translator 40 (or any other information other than the input text).

Additionally, claim 1 specifies a converter that converts the phonemes to additional FAP information and outputs the additional FAP information together with the FAP information separated by the decoder. No such converter exists in the Gasper system.

The Examiner asserted that sub-block 26 corresponds to the converter of claim1, but applicants respectfully disagree. Sub-block 26 is an audio generator, and the converter element of claim 1 does not even deal with audio matters.

Additionally still, claim 1 specifies a face rendering module that is response to an applied face model signal and to the output of the converter. The Examiner cited element 18, which is a video generator. The only teaching of what video generation unit 18 is, or does, is found in col. 5, lines 37-42, which state:

> The video output 19 and video generation 18 circuitry are controlled by the microprocessor 10 and share display RAM buffer space 22 to store and access memory mapped video. The video generation circuits also provide a sixty Hz timing signal interrupt to the microprocessor 10

This text implies that unit 18 is simply the conventional module of a computer that is responsible for creating the signals that are applied to the monitor, without any algorithmic processing . It does not suggest that unit 18 does any rendering (i.e., "express in another language or form; translate" *American Heritage Dictionary*, 1982) of a talking head, and it certainly does not describe any rendering from a "face model signal."

In view of the above analysis of the Gasper teachings, it is respectfully submitted that none of the element defined by claim 1 elements is anticipated, or made obvious, by the teachings of Gasper.

With reference to claim 12, the claim specifies

> a decoder, responsive to an input signal comprising signals representing audio and embedded video synthesis command signals, that separates the command signals from signals representing audio to develop an audio signal stream and a video synthesis command signals stream

Beutnagel 4-1-13-3

Based on the above analysis, it is clear that the Gasper system is not responsive to an "input signal comprising signals representing audio and embedded video synthesis command signals," that Gasper does not teach or suggest an element that separates such an input signal into video synthesis "command signals" and "signals representing audio," and does not create two streams as a result of such separating, where one stream is an "audio signal stream" and the other stream is a "video synthesis command signals stream." Therefore, it is respectfully submitted that claim 12 is also not anticipated by the Gasper reference. Since claim 12 is not anticipated by Gasper, it is respectfully submitted that claims 13-25 are also not anticipated by Gasper.

It is noted with references to claims 14 and 16, that the decoder is further specified to be one where, following separation of the video synthesis command signals from the input signals, the decoder converts text to elemental sound elements. As indicated above in connection with claim 1, it is quite clear that element 10 does not perform such action, because element 40 performs the translation from text to phonemes. Element 40, however, does no separating of video synthesis command signals from the input signals.

It is also noted with reference to claims 16-19, that FAPs are specified, and FAPs are neither described nor suggested by Gasper.

It is further noted that with reference to claim 20, that the converter defined in claim 16 not only is a speech synthesizer that is responsive to phoneme signals, but it also generates video synthesis command signals, and applies those generated command signals to the video synthesizer (specified in the base claim 12).

It is still further noted with reference to claims 21-25 that additional limitations are included regarding the structure of the element interconnections and the structure of the signals that are employed. All of the above-mentioned limitations serve to further distinguish the mentioned claims from the Gasper reference.

Claims 24 and 25 were rejected under 35 USC 103 as being unpatentable over of Gasper in view of Chen et al, US Patent 6,130,679. Applicants respectfully traverse.

The Chen reference is included for its teachings of FAP information. The Examiner asserted that Gasper teaches all of the limitations in claims 24 and 25, except

for the FAPs, which are taught by Chen et al. First, applicants have demonstrated above that the Examiner's first assertion is in error. Therefore, claims 24 and 25 are not obvious in view of the Gasper and Chen et al references. Second, claims 24 specify the use of particular FAPs. Chen et al do not teach that, and the Examiner does not assert that they do. Therefore, again, claims 24 and 25 are not obvious in view of the Gasper and Chen et al references.

Claims 29-30,41, and 42 were rejected under 35 USC 103 as being unpatentable over Goldenthal et al in view of Chen et al. The Examiner's reasoning in rejecting these claims parallels the Examiner's reasoning in rejecting claims 24 and 25. It is respectfully submitted that applicants' arguments regarding claims 24 and 25 apply with equal vigor to claims 29-30, 41 and 42.

In light of the above amendments and remarks, applicants respectfully submit that all of the Examiner's rejections have been overcome. Reconsideration and allowance of the outstanding claims are, respectfully, solicited.

Respectfully,
Mark Beutnagel
Ariel Fischer
Joern Ostermann
Yao Wang

Dated: 5/1/02

By _____
Henry T. Brendzel
Reg. No. 26,844
Phone (973) 467-2025
Fax (973) 467-6589
email brendzel@comcast.net

Beutnagel 4-1-13-3

Appendix showing changes made

In the Claims:

Delete claim 35.